

Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-01-24 14:39:35

PAGE 1

REFERENCE NO: 174

This contribution was submitted to the National Science Foundation as part of the NSF CI 2030 planning activity through an NSF Request for Information, https://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf17031. Consideration of this contribution in NSF's planning process and any NSF-provided public accessibility of this document does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government. The content of this submission is protected by the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>).

Author Names & Affiliations

- Felicito Jr Guerrero - US Dept. of Agriculture, Agricultural Research Service

Contact Email Address (for NSF use only)

(Hidden)

Research Domain, discipline, and sub-discipline

Genomics

Title of Submission

Computational Resource Needs for Agricultural Research on Non-Model Pests

Abstract (maximum ~200 words).

My research centers upon genomics of livestock pests and several of these pests have genome sizes that exceed that of the human genome. They also contain very large percentages of repetitive DNA, making assembly challenging.

Thus, I have been working with genome sequencing projects that require de novo assembly and long read technologies. I have run into difficulties with my own computational resources and that of academic collaborators because what we have available to us are just not powerful enough. We turned to TACC for assistance and they were able to provide access to enough computational power to enable us to error-correct large datasets of Pac Bio sequence reads and assemble the reads into a quality genome that is now available to the research community. We have embarked on two other similar projects and these will be greatly facilitated by access to TACC resources.

Question 1 Research Challenge(s) (maximum ~1200 words): Describe current or emerging science or engineering research challenge(s), providing context in terms of recent research activities and standing questions in the field.

My research centers upon genomics of livestock pests and several of these pests have genome sizes that exceed that of the human genome. They also contain very large percentages of repetitive DNA, making assembly challenging.

Thus, I have been working with genome sequencing projects that require de novo assembly and long read technologies. I have run into difficulties with my own computational resources and that of academic collaborators because what we have available to us are just not powerful enough. We turned to TACC for assistance and they were able to provide access to enough computational power to enable us to error-correct large datasets of Pac Bio sequence reads and assemble the reads into a quality genome that is now available to the research community. We have embarked on two other similar projects and these will be greatly facilitated by access to TACC resources.

Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-01-24 14:39:35

PAGE 2

REFERENCE NO: 174

Long read sequencing technologies utilized on non-model arthropod pests can be especially computationally intensive and access to computer power and algorithms for assembly of Pac Bio, Oxford Nanopore and coming long read technologies will be essential to continue the recent advances in agricultural genomics.

Question 2 Cyberinfrastructure Needed to Address the Research Challenge(s) (maximum ~1200 words): Describe any limitations or absence of existing cyberinfrastructure, and/or specific technical advancements in cyberinfrastructure (e.g. advanced computing, data infrastructure, software infrastructure, applications, networking, cybersecurity), that must be addressed to accomplish the identified research challenge(s).

I do not have what I would consider bold, forward-looking ideas, but I do want to point out the great usefulness of advanced computing systems to those of us working on non-model systems. The organisms I study have great economic impacts on livestock production systems around the world and having genome and transcriptome assembled sequences will advance studies to develop novel pest control technologies.

I urge you to continue to make these advanced computational resources available to as many researchers as possible. This will multiply the impacts of these resources and facilitate positive economic impacts to the agricultural sciences and agriculture in general.

Question 3 Other considerations (maximum ~1200 words, optional): Any other relevant aspects, such as organization, process, learning and workforce development, access, and sustainability, that need to be addressed; or any other issues that NSF should consider.

I wish to begin by thanking the staff at Texas Advanced Computing Center for assistance with integrating their resources with my genome sequencing project. My research centers upon genomics of livestock pests and several of these pests have genome sizes that exceed that of the human genome. They also contain very large percentages of repetitive DNA, making assembly challenging.

Thus, I have been working with genome sequencing projects that require de novo assembly and long read technologies. I have run into difficulties with my own computational resources and that of academic collaborators because what we have available to us are just not powerful enough. We turned to TACC for assistance and they were able to provide access to enough computational power to enable us to error-correct large datasets of Pac Bio sequence reads and assemble the reads into a quality genome that is now available to the research community. We have embarked on two other similar projects and these will be greatly facilitated by access to TACC resources.

I do not have what I would consider bold, forward-looking ideas, but I do want to point out the great usefulness of advanced computing systems to those of us working on non-model systems. The organisms I study have great economic impacts on livestock production systems around the world and having genome and transcriptome assembled sequences will advance studies to develop novel pest control technologies.

I urge you to continue to make these advanced computational resources available to as many researchers as possible. This will multiply the impacts of these resources and facilitate positive economic impacts to the agricultural sciences and agriculture in general.

Consent Statement

- "I hereby agree to give the National Science Foundation (NSF) the right to use this information for the purposes stated above and to display it on a publically available website, consistent with the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>)."

Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-01-24 14:39:35

REFERENCE NO: 174

PAGE 3

